

Integration of a Kalman filter in the geographically weighted regression for modeling the transmission of hand, foot and mouth disease

Bisong Hu (✉ hubisong624@126.com)

Jiangxi Normal University <https://orcid.org/0000-0003-3875-8792>

Wenqing Qiu

Jiangxi Normal University

Chengdong Xu

Institute of Geographic Sciences and Natural Resources Research Chinese Academy of Sciences

Jinfeng Wang

Institute of Geographic Sciences and Natural Resources Research Chinese Academy of Sciences

Research article

Keywords: Hand, foot and mouth disease, Kalman filter, Geographically weighted regression, Spatiotemporal pattern, Determinant factors

Posted Date: October 4th, 2019

DOI: <https://doi.org/10.21203/rs.2.15522/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at BMC Public Health on April 10th, 2020. See the published version at <https://doi.org/10.1186/s12889-020-08607-7>.

Abstract

Background: Hand, foot and mouth disease (HFMD) is a common infectious disease whose mechanism of transmission continues to remain a puzzle for researchers. The measurement and prediction of the HFMD incidence can be combined to improve the estimation accuracy, and provide a novel perspective to explore the spatiotemporal patterns and determinant factors of an HFMD epidemic.

Methods: In this study, we collected weekly HFMD incidence reports for a total of 138 districts in Shandong province, China, from May 2008 to March 2009. A Kalman filter was integrated with geographically weighted regression (GWR) to estimate the HFMD incidence. Spatiotemporal variation characteristics were explored and potential risk regions were identified, along with quantitatively evaluating the influence of meteorological and socioeconomic factors on the HFMD incidence.

Results: The results showed that the average error covariance of the estimated HFMD incidence by district was reduced from 0.3841 to 0.1846 compared to the measured incidence, indicating an overall improvement of over 50% in error reduction. Furthermore, three specific categories of potential risk regions of HFMD epidemics in Shandong were identified by the filter processing, with manifest filtering oscillations in the initial, local and long-term periods, respectively. Amongst meteorological and socioeconomic factors, the temperature and number of hospital beds per capita, respectively, were recognized as the dominant determinants that influence HFMD incidence variation.

Conclusions: The estimation accuracy of the HFMD incidence can be significantly improved by integrating a Kalman filter with GWR and the integration is effective for exploring spatiotemporal patterns and determinants of an HFMD epidemic. Our findings could help establish more accurate HFMD prevention and control strategies in Shandong. The present study demonstrates a novel approach to exploring spatiotemporal patterns and determinant factors of HFMD epidemics, and it can be easily extended to other regions and other infectious diseases similar to HFMD.

Background

Hand, foot and mouth disease (HFMD) is a common infectious disease caused by at least 20 enteroviruses including enterovirus 71 (EV-A71) and Coxsackie virus A16 (CA-V16) [1]. HFMD usually affects infants and children under five and its main symptoms include fever, mouth ulcers and blisters or vesicles on the hands, feet, and mouth. Existing vaccines are only partially effective for specific HFMD pathogens [2]. The transmission mechanism of HFMD epidemics is complicated and its spatiotemporal pattern is not yet fully understood [3]. During the last decades, HFMD has been widespread in Asian countries, such as Japan, Malaysia, and Singapore [4–6]. In China, the first large-scale outbreaks of HFMD occurred in Linyi city, Shandong province in 2007 [7] and in Fuyang city, Anhui province in 2008 [8]. Next, in May 2008 the Ministry of Health of China listed HFMD as a statutorily notifiable infectious category C disease. Numerous studies on HFMD epidemics were implemented in various regions,

particularly in provinces with serious epidemics, such as Guangdong [9, 10], Sichuan [11, 12], Henan [13, 14], Shandong [15, 16], and others.

Previous studies have mainly focused on characteristics of the epidemic [1, 13, 17], such as spatiotemporal patterns and correlations with various risk factors. HFMD epidemics have significant temporal variations and seasonality features, which vary between regions [18–21]. HFMD epidemics were spatially dispersed across counties in mainland China in the summer and winter, while clustered in spring and autumn; they were also geographically clustered in and closely linked to regions with high levels of monthly precipitation [3, 22]. In addition, HFMD epidemics follow complicated spatiotemporal patterns and transmission mechanisms, and are associated with several types of risk factors. For example, the HFMD incidence in Singapore has been found to be affected in a non-linear manner by the maximum temperature and rainfall, with a time lag of 1–2 weeks, and thresholds of 32 °C and 75 mm, respectively [23]. Furthermore, in Japan and Vietnam, temperature and humidity had significant effects on the HFMD incidence [19, 24]. The spatial variation of HFMD in counties across mainland China was found to be affected by a combination of climate variables, while the spatiotemporal transmission was largely driven by variations in temperature, with a 7-week lag [3]. Extreme precipitation was significantly associated with childhood HFMD in Hefei, China, and the susceptible risk in urban areas was much higher than that in rural ones [25]. High-risk areas of HFMD incidence temporally varied from northeast to southwest in Sichuan, China, and temperature and per capita gross domestic product (GDP) were the main positive driving factors [11].

Non-linear associations have been found between the HFMD incidence and meteorological, land-use, normalized difference vegetation index (NDVI) and socioeconomic factors in Shandong, China [16]. Many other studies have also focused on exploring of HFMD spatiotemporal patterns and the associated driving factors, by considering a variety of methods [3, 9–12, 14, 16, 18, 20–22, 24–32]. However, the measurement and prediction of the HFMD incidence are usually considered separately, and rarely in an integrated fashion. The former is mainly accomplished by using case reports, while the latter requires specific quantitative models. Considering both the measurement noise and the prediction uncertainty can positively improve the estimation accuracy of the HFMD incidence, and could possibly offer a fresh perspective in exploring spatiotemporal patterns and determinant factors of the epidemic. This study aims to estimate the spatiotemporal evolution of the HFMD incidence by districts using a Kalman filter integrated with geographically weighted regression (GWR), to explore the spatiotemporal variation characteristics and potential risk regions, and to quantitatively evaluate the influence of meteorological and socioeconomic factors on the HFMD variation.

Methods And Materials

Study region

Shandong is an eastern coastal province of China and is located between 34° 23' and 38° 24' north latitude and between 114° 48' and 122° 42' east longitude (Fig. 1). It extends to the Yellow Sea in the east

and is bordered by the Hebei, Henan, Anhui and Jiangsu provinces from northwest to southwest. The Shandong province has a total population of approximately 100.47 million and a total land area of 157,100 km². The gross domestic product (GDP) of Shandong province was 7,646.97 billion Yuan in 2018. Shandong falls in the warm temperate monsoon climate zone, with an annual average temperature and precipitation in the ranges of 11–14 °C and 550–590 mm, respectively. More than 60% of the annual rainfall in the Shandong province is registered in the summer, and high temperatures usually occur in seasons with high precipitation.

Data

From May 1st, 2008 to March 19th, 2009 (47 weeks), weekly HFMD incidence reports for a total of 138 districts in Shandong were collected from the Chinese Centre for Disease Control and Prevention. To reduce the influence of population size, weekly incidence rates were calculated to reflect the risk of the HFMD epidemic for sample locations, and the corresponding Thiessen polygons were constructed to account for spatial effects (Fig. 1). Monthly meteorological data from May 2008 to March 2009 were obtained from the China National Meteorological Information Center (<http://data.cma.cn/>), including the daily average, maximum, and minimum temperatures (°C), the air pressure (hPa), relative humidity (%), wind speed (m/s), precipitation (mm) and sunshine hours (h). The socioeconomic data were collected from the 2008 statistic Yearbook of Shandong province, including GDP (10,000 Yuan), ratio of the number of primary school students to the total population (%) and number of hospital beds per capita. Spatial Kriging methods were used to calculate the weekly average meteorological factors for each sample location during the 47-week study of HFMD epidemics. Both dynamic meteorological factors and static socioeconomic factors were normalized to the range of 0–1.

Geographically weighted regression model

Compared with the global multivariate regression model, local models can be more effective at describing potential local variations in relationships between dependent and independent variables. The geographically weighted regression [33, 34] is a typical local multivariate regression model extensively applied to measure the spatial relationships between variables and corresponding local variations across an entire area. Moreover, GWR model can clearly detect and interpret any non-stationary features of spatial patterns and associations, and has been widely used to estimate the epidemic risk and assess the influence of the epidemic determinants [35, 36]. The GWR model used in this study is as follows:

[Due to technical limitations, this equation is only available as a download in the supplemental files section.] (1)

where y_i is the HFMD incidence rate at location i with coordinates u_i and v_i , $\alpha(u_i, v_i)$ is the corresponding intercept constant, $x_{k,i}$ are a series of independent variables describing local variations, $\beta_k(u_i, v_i)$ are the

local regression coefficients to be estimated, which vary with location, $z_{i,j}$ are a series of independent variables connected with the global stability, γ_j are the corresponding stable coefficients, and ε_j indicates the estimation error.

To approximate the HFMD incidence rate of each sample location in Shandong province, we take the dynamic meteorological factors as the local variables x_k in the above GWR model, and the static socioeconomic factors as the global variables z_l . Therefore, every location in the study area has a set of specific coefficients to reflect the associations between the HFMD incidence rate and the global or local variables. To solve the proposed GWR model, we apply a Gaussian distance-decay function to represent the relative importance between locations and an adaptive kernel scheme to determine the bandwidth (optimal number of neighboring locations), which is calculated through an iterative optimization process according to the Akaike Information Criterion (AIC). Meanwhile, the significance of the estimated global/local coefficients was checked with pseudo t tests and the model significance was tested by variance analysis (F tests).

Kalman Filter

The Kalman filter (KF) is a data fusion algorithm initially designed to solve the discrete-data linear filtering problem and provides a recursive solution to estimate the state variable of a time-varying system [37, 38]. In this study, KF is used to estimate the HFMD incidence and quantitatively assess the influence of risk factors. For a specific district, we define a multivariate state space X , which includes the HFMD incidence and several static socioeconomic factors. The state space is time-varying and calculated using the following parametric formula:

[Due to technical limitations, this equation is only available as a download in the supplemental files section.] (2)

where X_t is the state vector containing the HFMD incidence and socioeconomic factors at time t , A is the state transition matrix indicating the effects of each state variable at time $t-1$ on the state vector at time t , U_t is a vector containing control variables which are dynamic meteorological and static socioeconomic factors relevant to this study, B is the control coefficient matrix indicating the effects of each control variable on the state vector, and ω_t is a random variable representing the process noise, which is drawn from a zero-mean Gaussian distribution $N(0, Q)$. Last, Q stands for the prediction noise variance and accounts for the prediction uncertainty compared with the real process. The prediction of the time-varying state vector could be implemented as follows:

[Due to technical limitations, this equation is only available as a download in the supplemental files section.] (3)

where \hat{X}_t is the prediction state vector at time t and X_{t-1} is the estimated (filtered) state vector at time $t-1$. The *a priori* estimation error covariance of the above prediction model propagates according to the equation:

[Due to technical limitations, this equation is only available as a download in the supplemental files section.] (4)

where P_t is the estimation error covariance of the prediction model at time t . Furthermore, by considering the HFMD incidence Y as the most important variable in the state vector X , we define a simple linear relationship linking the measurement Y to the state vector X :

[Due to technical limitations, this equation is only available as a download in the supplemental files section.] (5)

where Y_t is the measurement HFMD incidence at time t which is the observed incidence calculated based on the reported cases, C is the observation operator matrix, and v_t is a random variable representing the measurement noise which is also assumed to be drawn from a zero-mean Gaussian distribution $N(0, R)$. Similarly, R stands for the measurement noise variance and represents the measurement uncertainty.

When both the process prediction and the measurement are considered, the *a priori* estimation error covariance P_t and the measurement noise variance R are combined to generate the Kalman gain:

[Due to technical limitations, this equation is only available as a download in the supplemental files section.] (6)

where K_t stands for the Kalman gain at time t and is applied to compute the *a posteriori* estimation of the state vector at time t as the following linear combination of the *a priori* estimation \hat{X}_t and the actual measurement Y_t :

[Due to technical limitations, this equation is only available as a download in the supplemental files section.] (7)

As a function of the state vector covariance and the measurement noise, the Kalman gain K_t is noticeably high if the estimation error covariance is much higher than the measurement noise and the *a posteriori* estimation of the state vector significantly follows the measurements. Conversely, when K_t is low, the filter will essentially follow the predictions. In fact, K_t establishes the best combination between the process prediction and the measurement in order to minimize the mean square error between the *a posteriori* estimation X_t and its true value. After the update of the state vector as described above, the *a posteriori* estimation error covariance can be expressed as:

[Due to technical limitations, this equation is only available as a download in the supplemental files section.] (8)

where I is an identity matrix and P_t indicates the estimation error covariance after the prediction and the update at time t . The *a priori* estimations take place at each step of the recursive solution based on the last *a posteriori* estimations, according to Eqs. (3) and (4), the Kalman gain at each step is computed according to Eq. (6), and the *a posteriori* estimations which are also the *a priori* estimations of the next step are generated according to Eqs. (7) and (8). Beginning from the initial state, the prediction and the update appear at every single step of the KF recursive solution.

Integration of the Kalman filter with the GWR model

Weekly averages of HFMD incidences in the sample locations were collected; the corresponding spatial autocorrelation was weak, with a Moran's I of 0.0208 ($p = 0.5460$). However, the spatial stratified heterogeneity of the HFMD incidence among counties was statistically significant, with a GeoDetector q -statistic of 0.2153 ($p < 0.001$) [39, 40]. Therefore, GWR model was applied to explore the global or local associations between the HFMD incidence and meteorological or socioeconomic factors. The GWR model produced an overall coefficient of determination R^2 of 0.2182, which was only an approximately 14% improvement compared with the global regression prediction. These results were possibly caused by the measurement noise in the HFMD incidence, as well as the prediction noise of the GWR model. To better explore the spatiotemporal patterns and assess the determinant factors of the HFMD epidemic, we combined the Kalman filter with the GWR model. The filtering allows to couple the measured and predicted HFMD incidences, and improve the incidence estimation accuracy. On the other hand, GWR model indicates the associations between HFMD incidence and determinant factors, and therefore could provide the prediction modeling of state vector varying in the Kalman filter. Furthermore, the influence sensitivity of the control variables can be evaluated during the incidence filtering process, and the corresponding determinants of HFMD incidence can be quantitatively assessed.

In our proposed Kalman filter, the parameter C is a simple observation operator matrix that indicates the transition between the state vector X and the measured HFMD incidence Y . On the other hand, the state transition matrix A models the variation of the state vector that consists of the HFMD incidence and the static socioeconomic factors from time $t-1$ to time t , while the control coefficient matrix B describes the influence effects of the meteorological and socioeconomic factors on the state vector. Moreover, for different districts in the study area, the global and local effects of the determinant factors on the HFMD incidence vary spatially. Therefore, as shown in Fig. 2, we integrated the GWR model into the Kalman filter, derived the space-varying parameters A and B , and generated multiple filters for the various districts. The HFMD incidence was the explained variable in GWR model, as well as the measurement Y in the Kalman filter. The local and global explanatory variables in the GWR model were the meteorological and socioeconomic factors, which also constitutes the control vector U of the Kalman filter. Moreover, the state vector X in the Kalman filter contains the HFMD incidence and the socioeconomic factors. For each district, the global coefficients γ_l and the local coefficients $\beta_k(u_i, v_j)$, which indicate the associations between the HFMD incidence and determinant factors, were obtained from the GWR result. Thus, the

corresponding parameter A in the Kalman filter could be constructed from the global regression coefficients in the GWR model, while the parameter B using the local regression coefficients. Different from the parameters A and C , the control coefficient matrix B is district-dependent (various B s for districts), and the corresponding multiple filters describe the spatial variation of the HFMD incidence evolution patterns and determinant influence effects.

Results

Kalman filtering validation

The HFMD incidence rates of 138 monitored districts were obtained in 47 weeks (from May 1st, 2008 to March 19th, 2009). For each of the sample districts continuous weekly incidence rates were available, the week index varying from 1 to 47. The average weekly incidence by district varied with time and had a mean value of approximately 0.936×10^{-4} (in a range of 0.043×10^{-4} – 4.851×10^{-4}). Eight meteorological factors (air pressure, daily average, maximum, and minimum temperatures, precipitation, relative humidity, wind speed and sunshine hours) were selected as the local dynamic independent variables (u_1 – u_8), and the global static independent variables (u_9 – u_{11}) were the following three socioeconomic factors: GDP, ratio of primary school students and number of hospital beds per capita. Both dynamic and static variables were normalized to the range of 0–1.

To evaluate the overall efficiency of the Kalman filter for HFMD incidence assessment, weekly incidence rates and meteorological variables for the studied districts were first aggregated to weekly average values. Next, using the static socioeconomic variables, the regression coefficients were calculated with the ordinary least squares (OLS) linear regression method. Subsequently, these coefficients were applied to generate the parameters B and C within the Kalman filter model, and the initial prediction and measurement errors were assumed to be drawn from a standardized Gaussian distribution. As shown in Fig. 3–a, the filtering provided an adjustment to the weekly average HFMD incidences in the 138 districts to some extent compared to the corresponding measured values, and the estimated HFMD incidences followed a similar distribution as the measurements. Fig. 3–b) illustrates that the original measurement errors varied among districts, high-value errors correlating to districts with high-value measurements; the estimation errors after filtering only apparently approach zero (the blue error curve presents an approximately horizontal line around the x axis). Even in districts with high-value measured incidence, the Kalman filter satisfactorily reduces the estimation errors. The measurement and estimation errors of the HFMD incidences in districts are mapped in Fig. 4–a. The HFMD incidence errors were reduced from the range of -3.55×10^{-4} – 3.64×10^{-4} to -0.21×10^{-4} – 0.41×10^{-4} . The Kalman filter significantly reduced the incidence errors for the majority of the districts, especially for those with large measurement errors. Fig. 4–b illustrates the reduced error distribution after filtering: although several districts received negative error reductions, the errors that increased were small and approximately 10% of the reduced ones. Regions with large error reductions and large HFMD incidences had similar reduced error distributions and were surrounded by regions with negative error reductions (the light-yellow polygons surround the

dark-green ones). Overall, the Kalman filter plays an effective role for HFMD incidence assessment even if the filter parameters are derived from the OLS linear regression without spatial variances. The measurement error covariance was 0.5686, whereas the estimation error covariance was substantially reduced to 0.0211 after filtering.

The spatiotemporal pattern of HFMD incidence filtering

After the overall validation of the Kalman filtering for the HFMD incidence assessment, we applied this model to explore the spatiotemporal patterns of HFMD incidences for all 138 districts. The local and global coefficients of dynamic meteorological factors ($u_1 - u_8$) and static socioeconomic factors ($u_9 - u_{11}$) on the HFMD incidences of each district were separately calculated using the GWR model. The corresponding parameter B of the Kalman filter is a matrix array that includes 138 control coefficient matrices ($[11 \times 4]$), indicating the effects of meteorological and socioeconomic factors ($u_1 - u_{11}$) on the state vector for 138 districts, respectively. A total of 138 Kalman filters with spatial variations were used to assess the temporal changes of HFMD incidences in the studied districts under the determinant factors ($u_1 - u_{11}$). As shown in Fig. 5–a, the average errors of measured incidences started with a high initial value and varied from week 1 to week 47; the error interval of 1 standard deviation (1-StdDev) around the average showed local fluctuations, which are probably related to the abnormal temporal intervals of the HFMD incidence evolution. For instance, there was a tiny error increase that appeared in the 28th week (beginning on November 6th) accompanying a substantial interval expansion; the error intervals expanded significantly even when the error mean decreased to nearly zero in weeks 46 and 47 (beginning on March 12th). Fig. 5–b shows that, compared to the measurements, the error means and 1-StdDev intervals of measurement incidences were reduced. However, considering the above-mentioned temporal anomalies, even after filtering the error means and 1-StdDev intervals were still large in the first 8 weeks (beginning on May 1st). That is to say, the HFMD epidemic in Shandong probably had pronounced seasonality features, usually evolving from mid-March, increasing until late June and with a potential reversal in early November.

To explore the spatial variation of the HFMD incidence filtering, the error covariances of incidence measurements and estimations by district were analyzed as shown in Fig. 6–a. The majority of districts had satisfactory reductions of error covariances after filtering and several districts received noticeable reductions even when the original error covariances were large. However, the error covariances of several districts were still significant after filtering, and the Kalman filters played a weak role in these districts (their positions are indicated by red arrows). Fig. 6–b illustrates the spatial variation of the reduced error covariances by district after HFMD incidence filtering. The average error covariance of measured incidences was 0.3841, whereas the average estimated incidence error covariance was reduced to 0.1846, indicating an overall improvement of over 50% error reduction. However, several districts with significant error reductions overlapped to a certain extent with districts of large estimated incidence error

covariances (Fig. 6). In other words, the HFMD incidence evolutions in these districts were abnormal, deeming such areas as potential risk regions of HFMD epidemic outbreaks.

Further considerations were proposed in these specific districts, and among them (Fig. 6–b), error covariances of estimated incidences were classified in natural breaks and mapped in Fig. 7–a. Judging by the temporal variations of the filtered HFMD incidence errors in each district, three classes of potential risk areas were distinguished, and presented separately in Figs. 7–b, 7–c and 7–d, respectively. The temporal measurement curves and the estimation errors in two districts of the same class were extremely similar to each other. Although the spatial aggregation feature of these abnormal districts was weak (Fig. 7–a), we could still classify potential HFMD risk regions into three categories by using the Kalman filter model in association with the meteorological and socioeconomic factors. As shown in Fig. 7–b, the error curves of HFMD incidence filtering greatly varied in the early period but maintained a long-term steady trend. The second type of potential risk regions is illustrated in Fig. 7–c; such regions present a relatively long-term steady trend with slight variations within a few intervals. Last, the third type had significant oscillations during the long-term period and unsteady oscillations appear in unpredictable localized time intervals (Fig. 7–d). Evidently, the former two types of potential HFMD risk regions raise concerns during the localized periods, especially in HFMD high-incidence seasons. Although the risk regions of the latter type were probably characterized by relatively low incidences, the HFMD epidemic evolutions were unsteady in the long-term, thus more prevention and control policies (e.g. long-term epidemic surveillance) should be implemented in these specific districts. Overall, the proposed HFMD incidence filtering in Shandong showed a strong seasonal dependence and several specific potential HFMD risk regions were found without significant spatial clustering.

Influence sensitivity of determinant factors

The control coefficient matrix B of the Kalman filter was generated from the GWR results to indicate the relationships between the HFMD incidence and meteorological or socioeconomic factors. To assess the influence that each factor has on the HFMD incidence, we defined an index ζ_j ($j = 1-11$) to describe the assumed enhancement effect of determinant factors (u_1-u_{11}). Experiments were repeated to evaluate the influence sensitivity of each dynamic or static factor on the HFMD incidence filtering. In experiment j , ζ_j varied from 0 to 5 with a step of 0.5, which indicates that the enhancement effect of factor u_j had a step size of 50% increase, while the ζ_i ($i \neq j$) of other factors was kept invariant. The average errors and covariances of incidence estimations by district were applied to assess the influence sensitivity of meteorological and socioeconomic factors.

Figs. 8–a and 8–b demonstrate the variations of the average estimation errors and covariances of HFMD incidence filtering along with the variation of each meteorological factor. As expected, the temperature factors (u_2-u_4) played the most important roles in the relationship with HFMD incidence filtering, and the average estimation errors and covariances were both sensitive to their enhancement effects, suggesting

that higher temperature variations would cause a higher HFMD variation. Air pressure (u_1) was a secondary determinant affecting the HFMD variation approximately 25% as strongly as the temperature factors (Table 1). The next secondary determinants were sunshine hours, relative humidity, and precipitation (u_8, u_6, u_5). Compared to the latter rainfall factors, the effect of sunshine hours on the HFMD incidence variation was almost twice as much (Table 1). As shown in Figs. 8–a and 8–b, the wind speed (u_7) played a very weak role in HFMD incidence filtering, with a relative variation of nearly zero, reflecting that the HFMD epidemic probably had little airborne contagious transmission. Figs. 8–c and 8–d illustrate the influence of socioeconomic factors on the HFMD incidence filtering. The number of hospital beds per capita (u_{11}) was the dominant determinant, followed by the GDP (u_9), which influenced the HFMD incidence approximately 30% as strongly as the dominant factor (Table 1). The relative variation of HFMD incidence filtering with the ratio of primary school students (u_{10}) was very slight (Table 1), suggesting that the amount of susceptible population in the studied region was probably not the leading cause of the HFMD variation. Overall, the daily average, maximum, and minimum temperatures and air pressure were the dominant meteorological factors, while the number of hospital beds per capita and GDP were the dominant socioeconomic ones that influenced the HFMD incidence variation in Shandong. Concomitantly, the HFMD variation was extremely slight even at high values of the wind speed and ratio of primary school students.

Discussion

In recent years, Kalman filters have been extensively used in a variety of applications, such as land cover classification [41] or landslide susceptibility evaluation [42]. Typical applications in Earth science concentrate on remote sensing image processing [41, 43–46] and data assimilations in the fields of agriculture [47–49], agrology [50, 51], ecology [52], hydrology [53, 54], oceanography [55] and others. In epidemiology, Kalman filters are usually applied to the mathematical modeling of epidemic spreads for diseases such as HIV/AIDS and Ebola [56–58]. In the present study, a Kalman filter was used to estimate the spatiotemporal evolution of HFMD incidence in 138 districts of the Shandong province, China, by integration with a GWR model to identify the local relationships between the HFMD incidence and risk factors. The proposed integrated model showed significant improvement in the HFMD incidence estimation accuracy. The spatiotemporal variation characteristics and potential risk regions of HFMD incidence were explored, and the influence of meteorological and socioeconomic factors on the HFMD variation were assessed. The results showed that the Kalman filter was effective for the HFMD incidence assessment in Shandong and produced a reduction of error covariance from 0.5686 to 0.0211 at the provincial scale. Considering the spatial variation of Kalman filters for various districts, the error covariance was reduced from 0.3841 to 0.1846 after filtering. Furthermore, filter processing allowed to identify potential HFMD risk regions: three categories of risk regions could be distinguished, with manifest filtering oscillations in the initial, local and long-term periods, respectively. Although the detected potential risk regions did not exhibit significant spatial clustering, more attention should be paid to these districts, especially the ones in the third category, with long-term filtering oscillations.

In addition to exploring the HFMD spatiotemporal patterns, the influence sensitivity of meteorological and socioeconomic factors was determined. We found that three temperature factors were the dominant meteorological determinants of the HFMD epidemic in Shandong, although the air pressure also affected the HFMD epidemic to a certain extent; however, wind speed had no manifest effect. Intense variations of temperature or air pressure produced high variations of HFMD incidence, whereas the influence of wind speed on the epidemic incidence was negligible and unclear. Our findings are consistent with a number of previous studies [11, 16, 23, 24, 59, 60]. The environmental temperature relates to behavioral patterns such as increased contact among young children, thereby facilitating the spread of an HFMD infection [14]. However, our results indicate that meteorological factors such as precipitation, relative humidity and sunshine hours were not strongly associated with HFMD incidence, which is partially inconsistent with some of the previous studies. For instance, precipitation was strongly correlated with HFMD incidence in Singapore [23], and the number of HFMD cases increased significantly with increasing relative humidity in Japan [24]. HFMD cases at the county level across mainland China were spatially clustered and closely linked to the amounts of monthly precipitation in the region [22]. Relative humidity and precipitation were also found as the dominant driving factors of HFMD incidence in Henan, China [14]. Moreover, compared to GDP and ratio of primary school students to the total population, the number of hospital beds per capita appeared to be more dominant in HFMD incidence in Shandong. This result differs from other studies as well. For instance, GDP was the primary risk factor contributing to the spatial distribution of HFMD incidence in Sichuan and Henan, China [11, 14]. Possible reasons for this discrepancy include the differences between the studied regions, different transmission mechanisms of the HFMD epidemics, seasonal variations of meteorological factors, scale effects, zoning effects and others.

This study provides a multi-perspective on estimating the spread of an HFMD epidemic by combining measurement noise with prediction uncertainty and demonstrates a novel approach to exploring the spatiotemporal patterns and determinant factors of an HFMD epidemic. Nevertheless, there are several limitations to this study, described as follows. First, we generated the basic local associations between the HFMD incidence and meteorological and socioeconomic factors using a GWR model without considering an HFMD mathematical model. Also, a limited number of driving factors were selected, which could have led to an insufficient description and interpretation of the HFMD epidemic dynamic mechanism. Second, our method was trained on county-level data from the Shandong Province of China from 2008 to 2009, and applied only for the pattern exploration and risk assessment of the HFMD epidemic. This approach could easily be extended to other regions and infectious diseases similar to HFMD, although it should be accompanied by a thorough analysis and benchmarking of the model on the new problem. Lastly, it was hypothesized that the measurement and prediction noises of the Kalman filter followed a zero-mean Gaussian distribution, and the model of the state vector and control variables was linear. These assumptions might have limited the applicability of the model; appropriate improvements could include non-linear filters and non-Gaussian noise distributions such as an extended Kalman filter (EKF), an unscented Kalman filter (UKF), or a particle filter (PF).

Conclusion

This study introduces a novel perspective to explore the spatiotemporal patterns and determinant factors of an HFMD epidemic. To this purpose, a Kalman filter method integrated with the GWR model with the aim to identify the global and local relationships between HFMD incidence and dynamic meteorological and static socioeconomic factors was designed. The proposed method considers both measurement noise and prediction uncertainty, which reduces the estimation error covariance of the HFMD incidence and improves the estimation accuracy. The filter processing could help explore the spatiotemporal patterns and determinants of the HFMD epidemic. As a result, three specific categories of potential risk regions of HFMD epidemics in Shandong were identified, with temperature factors and number of hospital beds per capita as the dominant determinants of the epidemic incidence. Furthermore, our approach can be extended to other regions and other infectious diseases similar to HFMD.

Abbreviations

CDC: Chinese Center for Disease Control and Prevention; EKF: Extended Kalman filter; GDP: Gross domestic product; GWR: Geographically weighted regression; HFMD: Hand, foot and mouth disease; KF: Kalman filter; NDVI: Normalized difference vegetation index; OLS: Ordinary least squares; PF: Particle filter; StdDev: Standard deviation; UKF: Unscented Kalman filter.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The meteorological data used in the study are available from the China National Meteorological Information Center (<http://data.cma.cn/>). The Socioeconomic data used in the study are available from the statistical Yearbook of Shandong province (<http://tjj.shandong.gov.cn/>). For the other data, please contact the authors for a link to the raw data.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported by the following grants: National Natural Science Foundation of China (41531179, 41421001, 41961055); National Key R&D Program of China (2016YFC1302504); Innovation Project of LREIS (O88RA200YA); Opening Fund of Key Laboratory of Poyang Lake Wetland and Watershed Research (Jiangxi Normal University), Ministry of Education (No. PK2019001).

Authors' contributions

BH completed the design of the study, implemented the model integration, finished the data analysis and interpretation, and wrote the article. WQ conducted the data pretreatment and analysis, conducted the GWR and GeoDetector computing. CX prepared the HFMD and meteorological or socioeconomic data and participated in the analysis of the results and in the writing. JW participated in the modeling design and in the writing and accomplished the manuscript revision. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

Authors' information

Not applicable.

References

1. Xing W, Liao Q, Viboud C, Zhang J, Sun J, Wu JT, et al. Hand, foot, and mouth disease in China, 2008–12: an epidemiological study. *Lancet Infect Dis*. 2014;14:308–18.
2. Zhu F-C, Meng F-Y, Li J-X, Li X-L, Mao Q-Y, Tao H, et al. Efficacy, safety, and immunology of an inactivated alum-adjuvant enterovirus 71 vaccine in children in China: a multicentre, randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet*. 2013;381:2024–32.
3. Wang J-F, Xu C-D, Tong S-L, Chen H-Y, Yang W-Z. Spatial dynamic patterns of hand-foot-mouth disease in the People's Republic of China. *Geospatial Health*. 2013;7:381–90.
4. Ang LW, Koh BKW, Chan KP, Chua LT, James L, Goh KT. Epidemiology and Control of Hand, Foot and Mouth Disease in Singapore, 2001–2007. *Ann Acad Med Singap*. 2009;38:106–12.
5. Chua KB, Kasri AR. Hand foot and mouth disease due to enterovirus 71 in Malaysia. *Virolog Sin*. 2011;26:221.

6. Hosoya M, Kawasaki Y, Sato M, Honzumi K, Hayashi A, Hiroshima T, et al. Genetic diversity of coxsackievirus A16 associated with hand, foot, and mouth disease epidemics in Japan from 1983 to 2003. *J Clin Microbiol.* 2007;45:112–20.
7. Zhang Y, Tan X-J, Wang H-Y, Yan D-M, Zhu S-L, Wang D-Y, et al. An outbreak of hand, foot, and mouth disease associated with subgenotype C4 of human enterovirus 71 in Shandong, China. *Journal of Clinical Virology.* 2009;44:262–7.
8. Zhang Y, Zhu Z, Yang W, Ren J, Tan X, Wang Y, et al. An emerging recombinant human enterovirus 71 responsible for the 2008 outbreak of Hand Foot and Mouth Disease in Fuyang city of China. *Virology Journal.* 2010;7:94.
9. Guo C, Yang J, Guo Y, Ou Q-Q, Shen S-Q, Ou C-Q, et al. Short-term effects of meteorological factors on pediatric hand, foot, and mouth disease in Guangdong, China: a multi-city time-series analysis. *BMC Infect Dis.* 2016;16:524.
10. Du Z, Lawrence WR, Zhang W, Zhang D, Yu S, Hao Y. Interactions between climate factors and air pollution on daily HFMD cases: A time series study in Guangdong, China. *Sci Total Environ.* 2019;656:1358–64.
11. Liao J, Qin Z, Zuo Z, Yu S, Zhang J. Spatial-temporal mapping of hand foot and mouth disease and the long-term effects associated with climate and socio-economic variables in Sichuan Province, China from 2009 to 2013. *Sci Total Environ.* 2016;563:152–9.
12. Song C, Shi X, Bo Y, Wang J, Wang Y, Huang D. Exploring spatiotemporal nonstationary effects of climate factors on hand, foot, and mouth disease using Bayesian Spatiotemporally Varying Coefficients (STVC) model in Sichuan, China. *Science of The Total Environment.* 2019;648:550–60.
13. Huang X, Wei H, Wu S, Du Y, Liu L, Su J, et al. Epidemiological and Etiological Characteristics of Hand, Foot, and Mouth Disease in Henan, China, 2008–2013. *Sci Rep.* 2015;5:8904.
14. Xu C, Zhang X, Xiao G. Spatiotemporal decomposition and risk determinants of hand, foot and mouth disease in Henan, China. *Sci Total Environ.* 2019;657:509–16.
15. Wang J, Hu T, Sun D, Ding S, Carr MJ, Xing W, et al. Epidemiological characteristics of hand, foot, and mouth disease in Shandong, China, 2009–2016. *Sci Rep.* 2017;7:8900.
16. Li L, Qiu W, Xu C, Wang J. A spatiotemporal mixed model to assess the influence of environmental and socioeconomic factors on the incidence of hand, foot and mouth disease. *BMC Public Health.* 2018;18:274.
17. Wang Y, Feng Z, Yang Y, Self S, Gao Y, Longini IM, et al. Hand, Foot, and Mouth Disease in China Patterns of Spread and Transmissibility. *Epidemiology.* 2011;22:781–92.
18. Liu W, Ji H, Shan J, Bao J, Sun Y, Li J, et al. Spatiotemporal Dynamics of Hand-Foot-Mouth Disease and Its Relationship with Meteorological Factors in Jiangsu Province, China. *PLoS One.* 2015;10:e0131311.
19. Phung D, Nguyen HX, Nguyen HLT, Do CM, Tran QD, Chu C. Spatiotemporal variation of hand-foot-mouth disease in relation to socioecological factors: A multiple-province analysis in Vietnam. *Sci Total Environ.* 2018;610:983–91.

20. Zhang X, Xu C, Xiao G. Space-time heterogeneity of hand, foot and mouth disease in children and its potential driving factors in Henan, China. *BMC Infect Dis.* 2018;18:638.
21. Zhao J, Hu X. The complex transmission seasonality of hand, foot, and mouth disease and its driving factors. *BMC Infect Dis.* 2019;19:521.
22. Wang J, Guo Y-S, Christakos G, Yang W-Z, Liao Y-L, Li Z-J, et al. Hand, foot and mouth disease: spatiotemporal transmission and climate. *Int J Health Geogr.* 2011;10:25.
23. Hii YL, Rocklov J, Ng N. Short Term Effects of Weather on Hand, Foot and Mouth Disease. *PLoS One.* 2011;6:e16796.
24. Onozuka D, Hashizume M. The influence of temperature and humidity on the incidence of hand, foot, and mouth disease in Japan. *Sci Total Environ.* 2011;410:119–25.
25. Cheng J, Wu J, Xu Z, Zhu R, Wang X, Li K, et al. Associations between extreme precipitation and childhood hand, foot and mouth disease in urban and rural areas in Hefei, China. *Sci Total Environ.* 2014;497:484–90.
26. Xie Y, Chongsuvivatwong V, Tang Z, McNeil EB, Tan Y. Spatio-Temporal Clustering of Hand, Foot, and Mouth Disease at the County Level in Guangxi, China. *PLoS One.* 2014;9:e88065.
27. Shi RX, Wang JF, Xu CD, Lai SJ, Yang WZ. Spatiotemporal pattern of hand-foot-mouth disease in China: an analysis of empirical orthogonal functions. *Public Health.* 2014;128:367–75.
28. Gui J, Liu Z, Zhang T, Hua Q, Jiang Z, Chen B, et al. Epidemiological Characteristics and Spatial-Temporal Clusters of Hand, Foot, and Mouth Disease in Zhejiang Province, China, 2008–2012. *PLoS One.* 2015;10:e0139109.
29. Wang Y, Lai Y, Du Z, Zhang W, Feng C, Li R, et al. Spatiotemporal Distribution of Hand, Foot, and Mouth Disease in Guangdong Province, China and Potential Predictors, 2009–2012. *Int J Environ Res Public Health.* 2019;16:1191.
30. Yu G, Li Y, Cai J, Yu D, Tang J, Zhai W, et al. Short-term effects of meteorological factors and air pollution on childhood hand-foot-mouth disease in Guilin, China. *Sci Total Environ.* 2019;646:460–70.
31. Huang R, Ning H, He T, Bian G, Hu J, Xu G. Impact of PM10 and meteorological factors on the incidence of hand, foot, and mouth disease in female children in Ningbo, China: a spatiotemporal and time-series study. *Environ Sci Pollut Res.* 2019;26:17974–85.
32. Zhang Q, Zhou M, Yang Y, You E, Wu J, Zhang W, et al. Short-term effects of extreme meteorological factors on childhood hand, foot, and mouth disease reinfection in Hefei, China: A distributed lag non-linear analysis. *Sci Total Environ.* 2019;653:839–48.
33. Brunson C, Fotheringham AS, Charlton ME. Geographically weighted regression: A method for exploring spatial nonstationarity. *Geogr Anal.* 1996;28:281–98.
34. Fotheringham AS, Charlton ME, Brunson C. Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. *Environ Plan A.* 1998;30:1905–27.

35. Hong Z, Hao H, Li C, Du W, Wei L, Wang H. Exploration of potential risks of Hand, Foot, and Mouth Disease in Inner Mongolia Autonomous Region, China Using Geographically Weighted Regression Model. *Sci Rep.* 2018;8:17707.
36. Hu M, Li Z, Wang J, Jia L, Liao Y, Lai S, et al. Determinants of the Incidence of Hand, Foot and Mouth Disease in China Using Geographically Weighted Regression Models. *PLoS One.* 2012;7:e38978.
37. Kalman RE. A new approach to linear filtering and prediction problems. *J Basic Eng.* 1960;82:35–45.
38. Kalman RE, Bucy RS. New results in linear filtering and prediction theory. *J Basic Eng.* 1961;83:95–108.
39. Wang J-F, Li X-H, Christakos G, Liao Y-L, Zhang T, Gu X, et al. Geographical Detectors-Based Health Risk Assessment and its Application in the Neural Tube Defects Study of the Heshun Region, China. *International Journal of Geographical Information Science.* 2010;24:107–27.
40. Wang J-F, Zhang T-L, Fu B-J. A measure of spatial stratified heterogeneity. *Ecological Indicators.* 2016;67:250–6.
41. Kleynhans W, Olivier JC, Wessels KJ, van den Bergh F, Salmon BP, Steenkamp KC. Improving Land Cover Class Separation Using an Extended Kalman Filter on MODIS NDVI Time-Series Data. *IEEE Geosci Remote Sens Lett.* 2010;7:381–5.
42. Gorsevski PV, Jankowski P. An optimized solution of multi-criteria evaluation analysis of landslide susceptibility using fuzzy sets and Kalman filter. *Computers & Geosciences.* 2010;36:1005–20.
43. Samain O, Roujean J-L, Geiger B. Use of a Kalman filter for the retrieval of surface BRDF coefficients with a time-evolving model based on the ECOCLIMAP land cover classification. *Remote Sensing of Environment.* 2008;112:1337–46.
44. Garzelli A, Nencini F. Panchromatic sharpening of remote sensing images using a multiscale Kalman filter. *Pattern Recognition.* 2007;40:3568–77.
45. Salmon BP, Kleynhans W, Olivier JC, van den Bergh F, Wessels KJ. A modified temporal criterion to meta-optimize the extended Kalman filter for land cover classification of remotely sensed time series. *International Journal of Applied Earth Observation and Geoinformation.* 2018;67:20–9.
46. Kanakaraj S, Nair MS, Kalady S. Adaptive Importance Sampling Unscented Kalman Filter based SAR image super resolution. *Computers & Geosciences.* 2019;133:104310.
47. Li R, Li C, Dong Y, Liu F, Wang J, Yang X, et al. Assimilation of Remote Sensing and Crop Model for LAI Estimation Based on Ensemble Kalman Filter. *Agricultural Sciences in China.* 2011;10:1595–602.
48. Zhao Y, Chen S, Shen S. Assimilating remote sensing information with crop model using Ensemble Kalman Filter for improving LAI monitoring and yield estimation. *Ecological Modelling.* 2013;270:30–42.
49. Huang J, Sedano F, Huang Y, Ma H, Li X, Liang S, et al. Assimilating a synthetic Kalman filter leaf area index series into the WOFOST model to improve regional winter wheat yield estimation. *Agricultural and Forest Meteorology.* 2016;216:188–202.

50. Huang C, Li X, Lu L, Gu J. Experiments of one-dimensional soil moisture assimilation system based on ensemble Kalman filter. *Remote Sensing of Environment*. 2008;112:888–900.
51. Gruber A, De Lannoy G, Crow W. A Monte Carlo based adaptive Kalman filtering framework for soil moisture data assimilation. *Remote Sensing of Environment*. 2019;228:105–14.
52. Chen C, Huang J, Chen Q, Zhang J, Li Z, Lin Y. Assimilating multi-source data into a three-dimensional hydro-ecological dynamics model using Ensemble Kalman Filter. *Environmental Modelling & Software*. 2019;117:188–99.
53. Xie X, Zhang D. Data assimilation for distributed hydrological catchment modeling via ensemble Kalman filter. *Advances in Water Resources*. 2010;33:678–90.
54. Zou L, Zhan C, Xia J, Wang T, Gippel CJ. Implementation of evapotranspiration data assimilation with catchment scale distributed hydrological model via an ensemble Kalman Filter. *Journal of Hydrology*. 2017;549:685–702.
55. Shu Y, Zhu J, Wang D, Xiao X. Assimilating remote sensing and in situ observations into a coastal model of northern South China Sea using ensemble Kalman filter. *Continental Shelf Research*. 2011;31:S24–36.
56. Cazelles B, Chau NP. Using the Kalman filter and dynamic models to assess the changing HIV/AIDS epidemic. *Mathematical Biosciences*. 1997;140:131–54.
57. Cobb L, Krishnamurthy A, Mandel J, Beezley JD. Bayesian tracking of emerging epidemics using ensemble optimal statistical interpolation. *Spatial and Spatio-temporal Epidemiology*. 2014;10:39–48.
58. Ndanguza D, Mbalawata IS, Haario H, Tchuente JM. Analysis of bias in an Ebola epidemic model by extended Kalman filter approach. *Mathematics and Computers in Simulation*. 2017;142:113–29.
59. Chang H-L, Chio C-P, Su H-J, Liao C-M, Lin C-Y, Shau W-Y, et al. The Association between Enterovirus 71 Infections and Meteorological Parameters in Taiwan. *PLoS One*. 2012;7:e46845.
60. Wei J, Hansen A, Liu Q, Sun Y, Weinstein P, Bi P. The Effect of Meteorological Variables on the Transmission of Hand, Foot and Mouth Disease in Four Major Cities of Shanxi Province, China: A Time Series Data Analysis (2009–2013). *Plos Neglect Trop Dis*. 2015;9:e0003572.

Table 1

Table 1 Relative variations of average errors and covariances of HFMD incidence filtering with meteorological and socioeconomic factors.

Determinant factors	Variable	Relative variation (%)	
		Average error	Average covariance
Air pressure (hPa)	u_1	4.95	2.78
Daily average temperature (°C)	u_2	21.05	11.92
Daily maximum temperature (°C)	u_3	18.05	10.65
Daily minimum temperature (°C)	u_4	17.14	8.07
Precipitation (mm)	u_5	1.48	0.48
Relative humidity (%)	u_6	1.49	0.42
Wind speed (m/s)	u_7	0.44	0.12
Sunshine hours (h)	u_8	2.77	0.96
Gross domestic product (GDP) (10^4 CNY)	u_9	0.65	1.54
Ratio of primary school students (%)	u_{10}	0.19	0.44
Number of hospital beds per capita	u_{11}	2.78	4.95

Figures

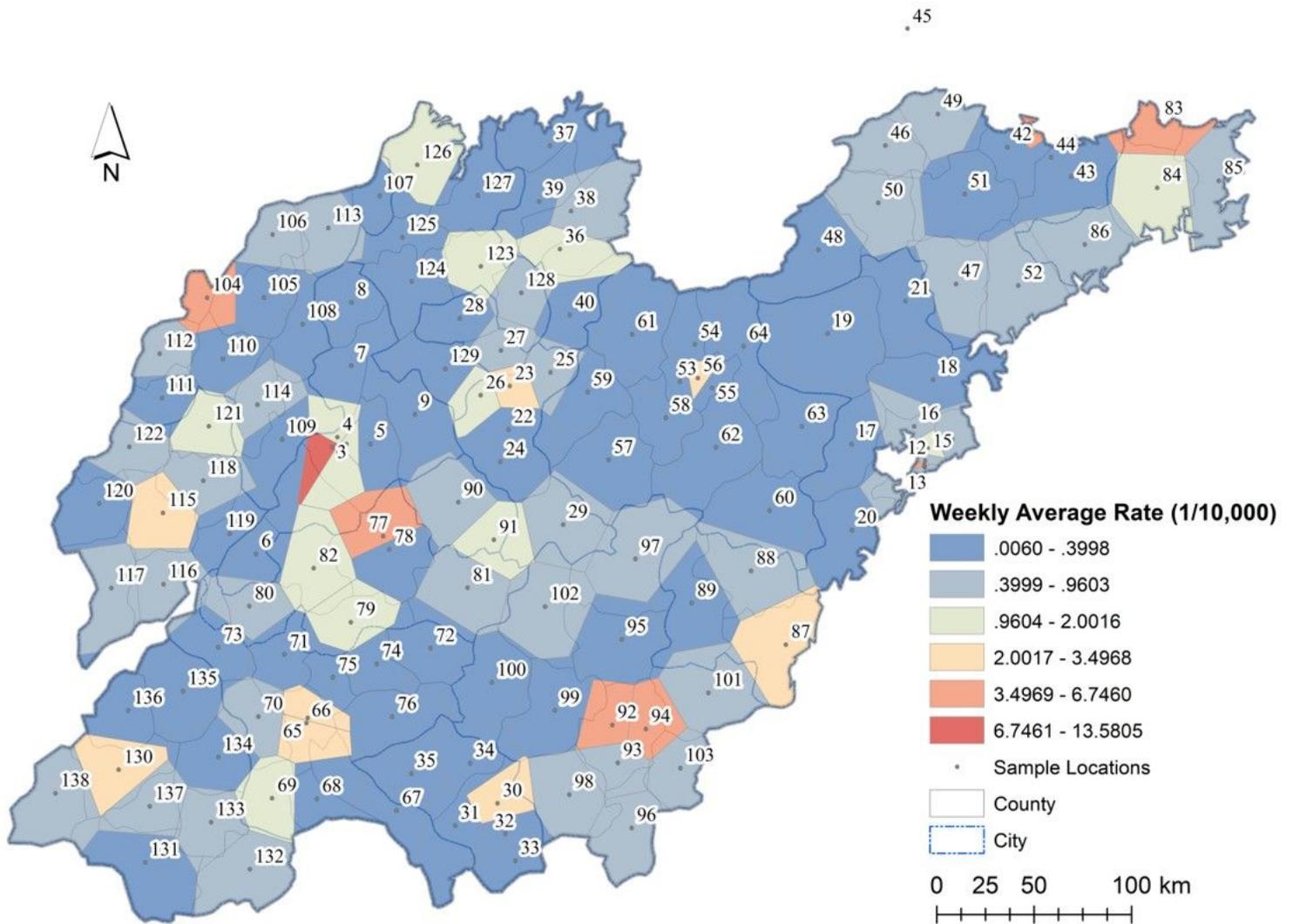


Figure 1

Location of Shandong province and weekly average incidence rates of HFMD in Thiessen polygons with 138 central sample locations. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

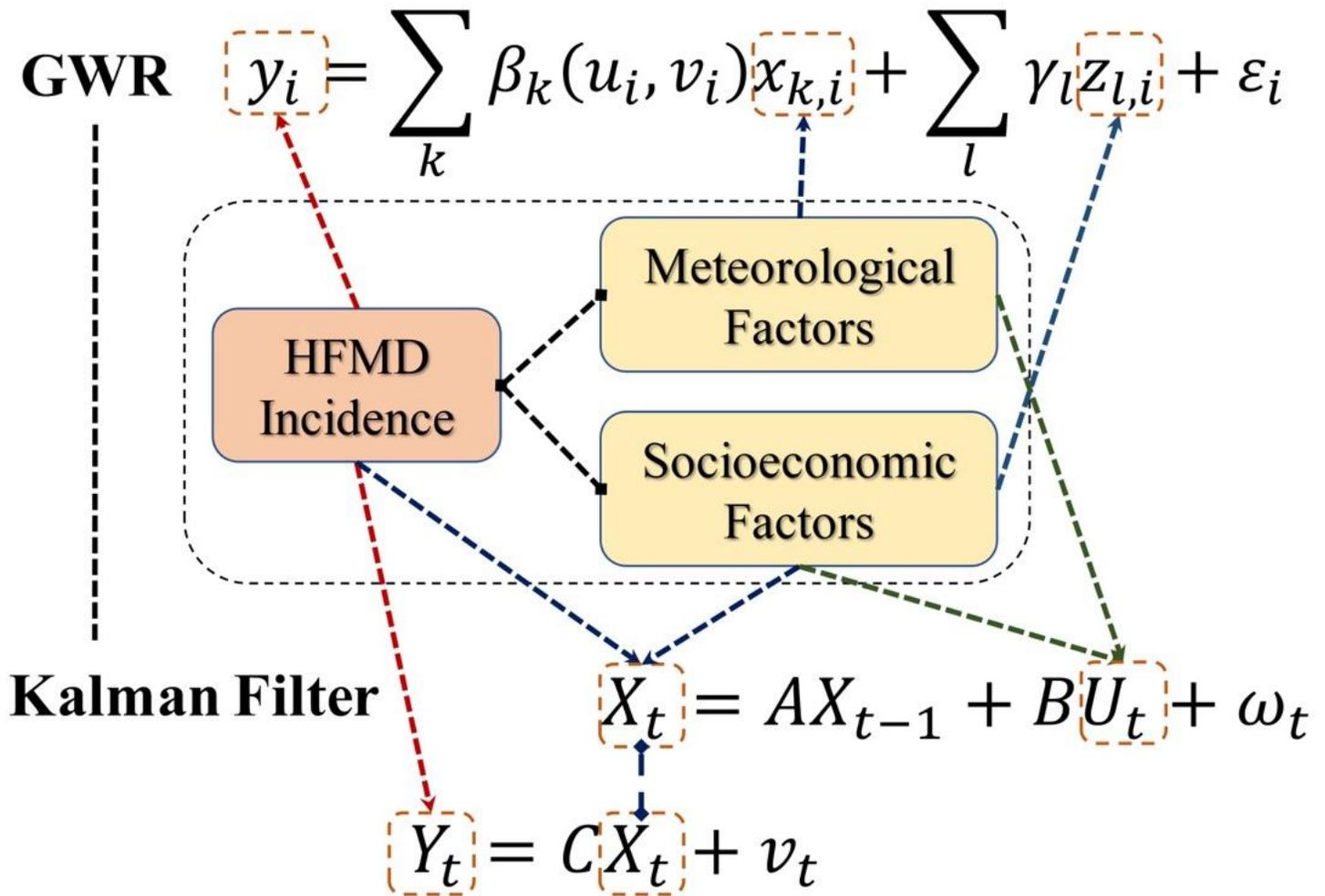


Figure 2

Integration framework of the Kalman Filter with the GWR model.

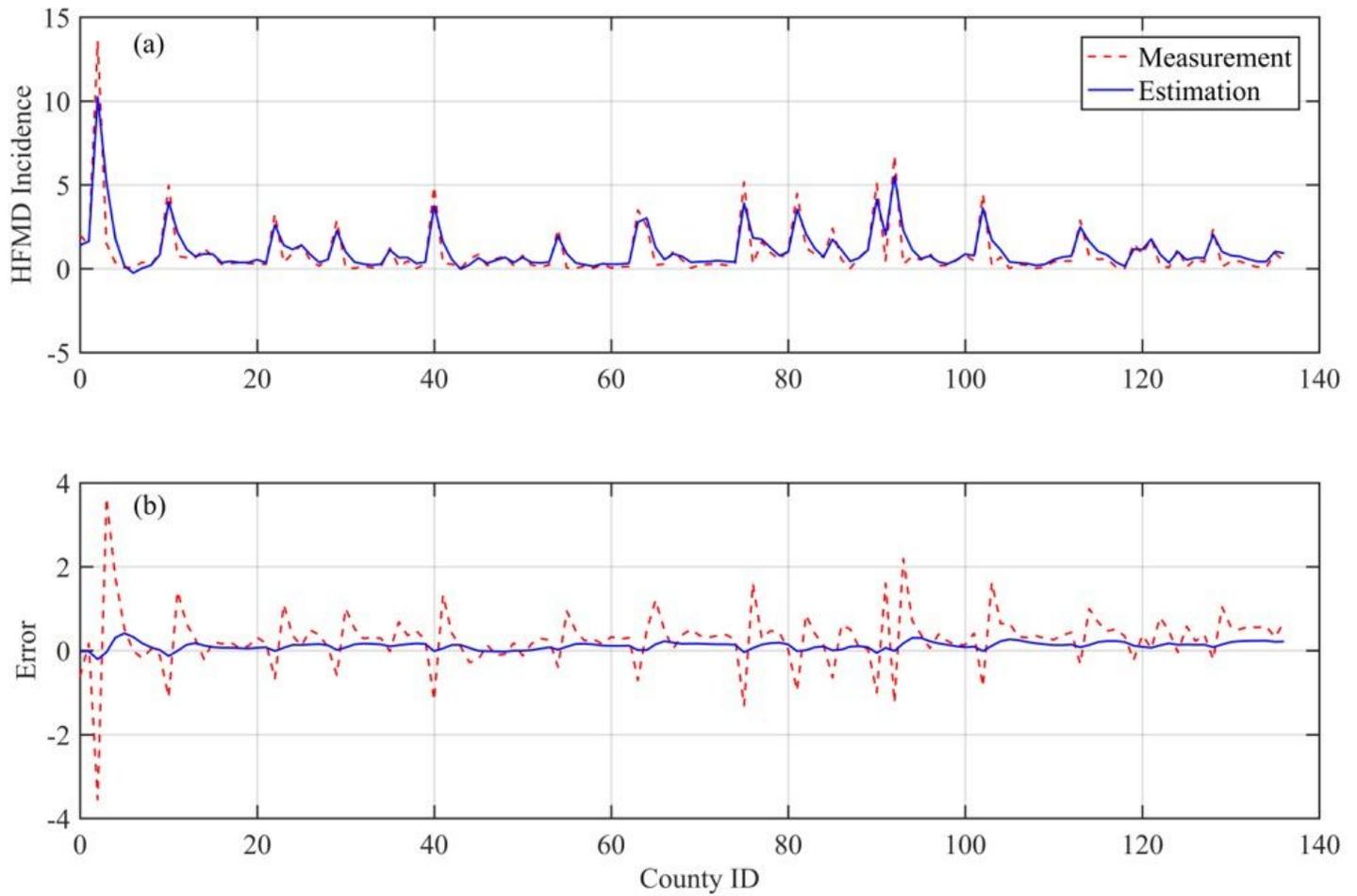


Figure 3

Efficiency evaluation of the Kalman Filter for the HFMD incidence assessment. (a) Measurement and estimation incidences. (b) Measurement and estimation errors.

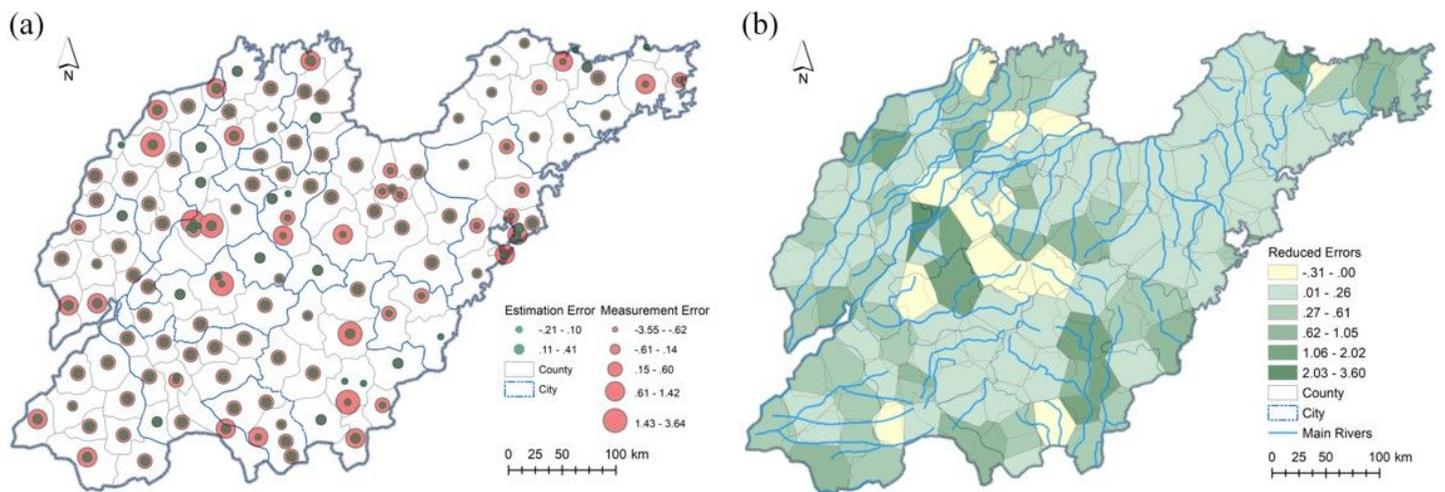


Figure 4

Error distributions of the global HFMD incidence filtering. (a) Measurement and estimation errors. (b) Reduced errors. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

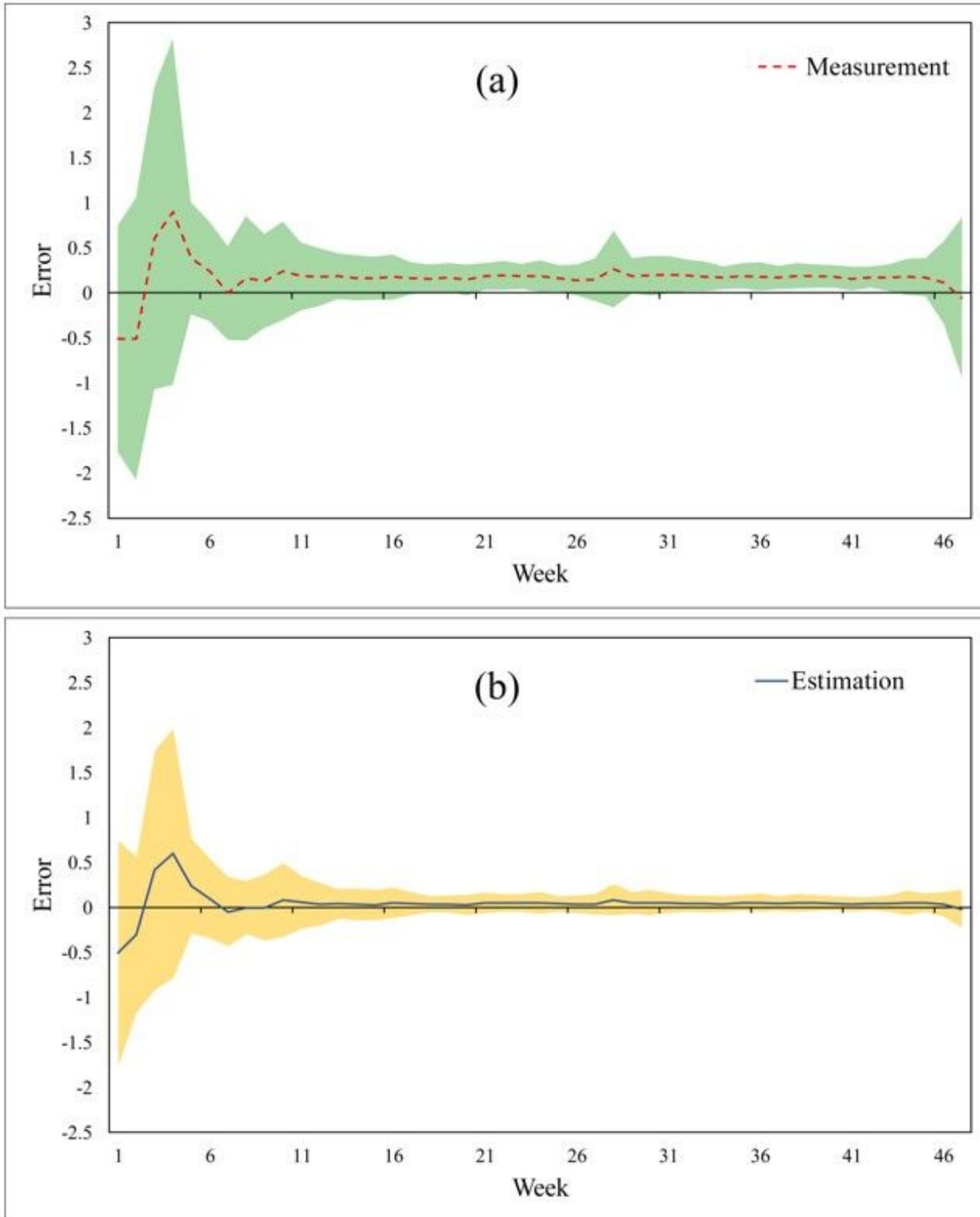


Figure 5

Temporal variation of the average errors of filtered HFMD incidence by district (the shaded area denotes the interval of 1 standard deviation around the average error). (a) Measurement errors. (b) Estimation errors.

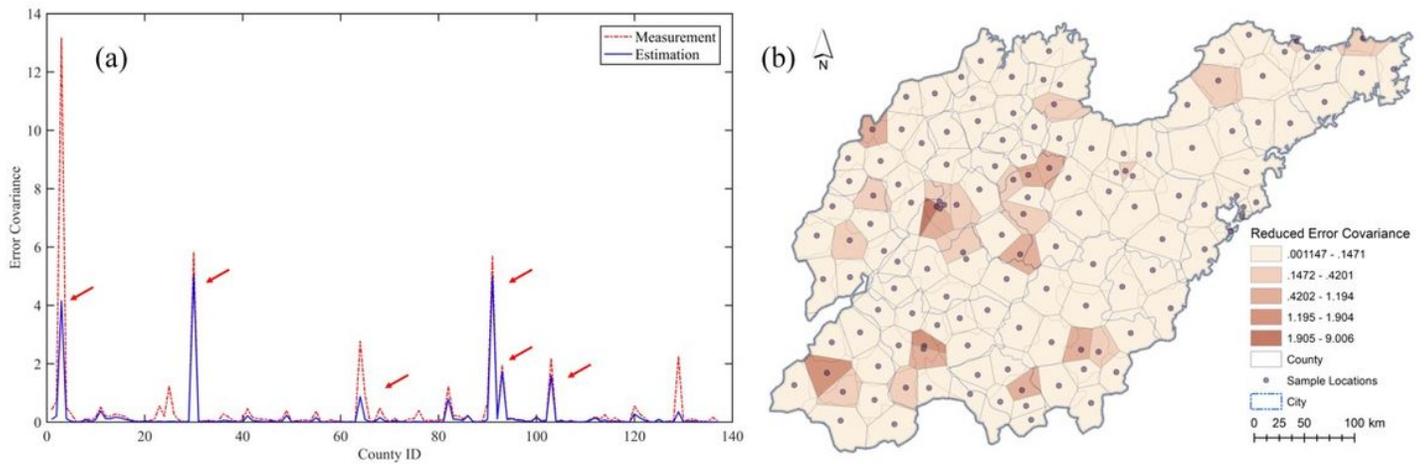


Figure 6

Spatial variation of error covariances of HFMD incidence filtering in districts. (a) Measurement and estimation covariances. (b) Reduced covariance distribution. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

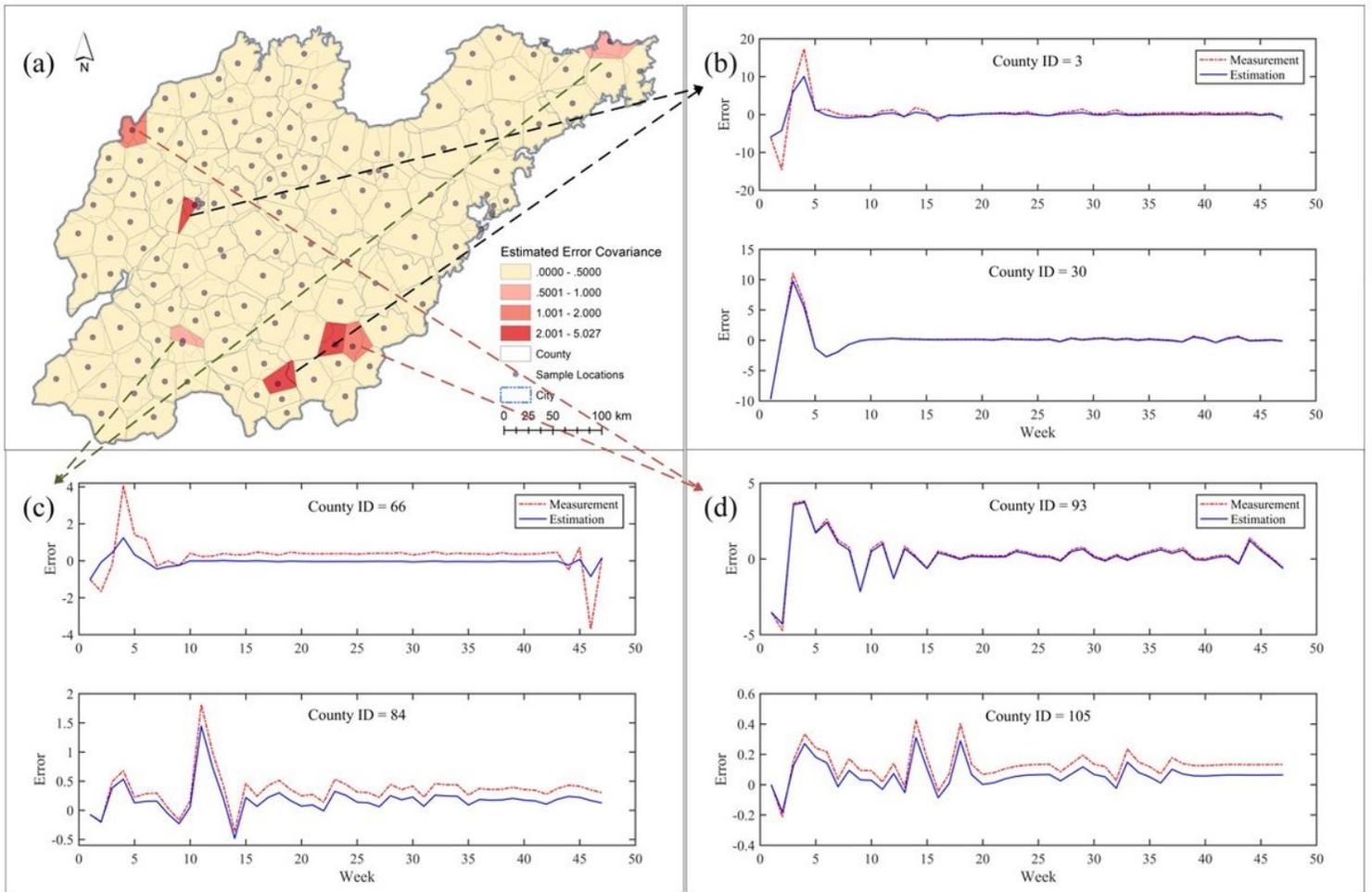


Figure 7

Potential HFMD risk regions and the separate error variations of incidence filtering. (a) Distribution of risk regions. Error variations of incidence filtering of (b) 1st category; (c) 2nd category; (d) 3rd category. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

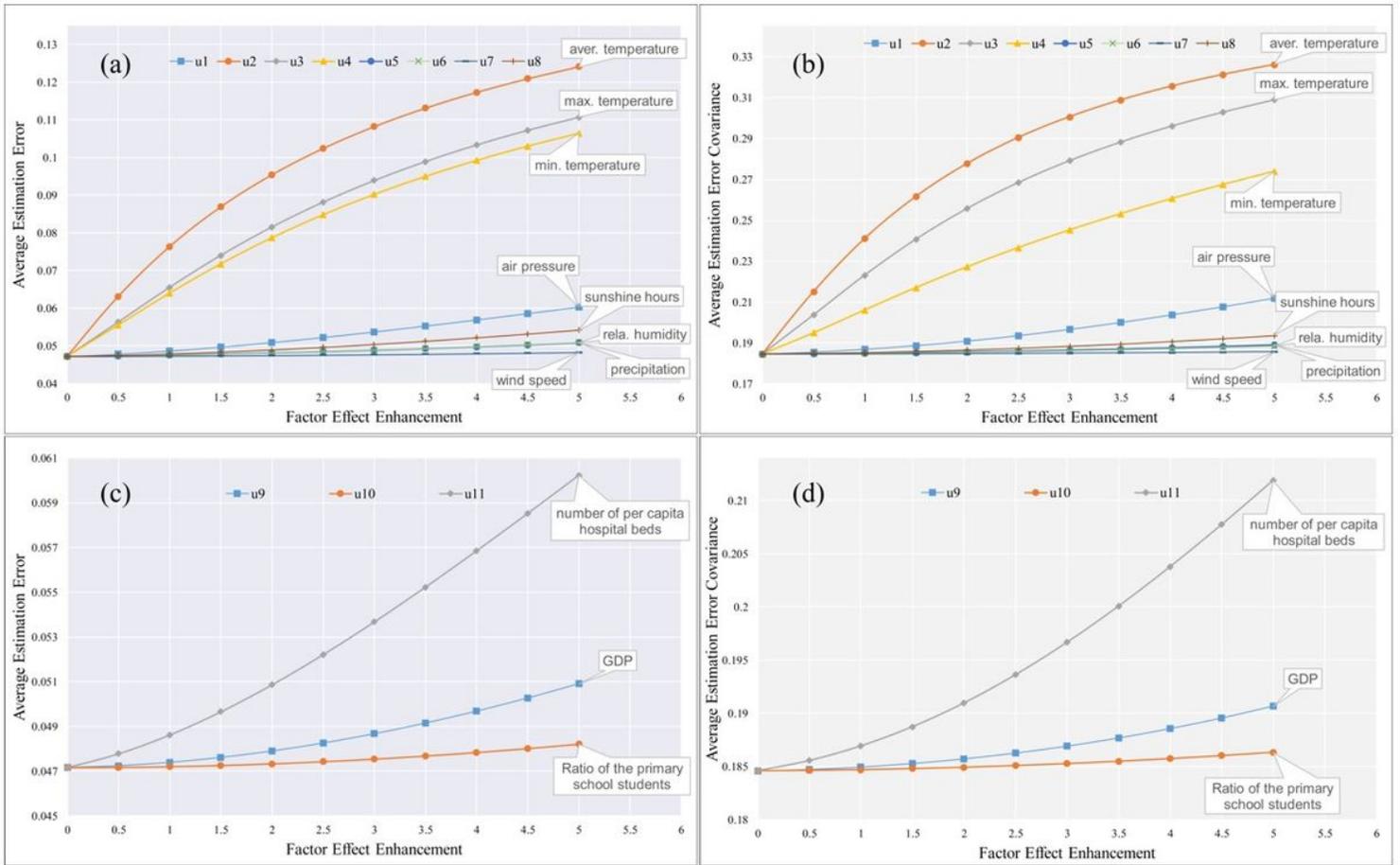


Figure 8

Influence sensitivity of HFMD incidence filtering along with meteorological and socioeconomic factors. (a) Average errors and (b) covariances of meteorological factors (u1–u8). (c) Average errors and (d) covariances of socioeconomic factors (u9–u11).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [eq5.jpg](#)
- [eq3.jpg](#)
- [eq4.jpg](#)
- [eq2.jpg](#)
- [eq8.jpg](#)
- [eq7.jpg](#)
- [eq6.jpg](#)
- [eq1.jpg](#)